

DOCUMENT RESUME

ED 469 291

EC 309 219

AUTHOR Grossen, Bonnie; Hagen-Burke, Shanna; Burke, Mack D.
TITLE An Experimental Study of the Effects of Considerate Curricula in Language Arts on Reading Comprehension and Writing. Research Report.
INSTITUTION Kansas Univ., Lawrence. Inst. for Academic Access.
SPONS AGENCY Special Education Programs (ED/OSERS), Washington, DC.
REPORT NO RR-13
PUB DATE 2002-00-00
NOTE 56p.
CONTRACT 84.324S
PUB TYPE Reports - Research (143)
EDRS PRICE EDRS Price MF01/PC03 Plus Postage.
DESCRIPTORS *Curriculum Design; *Disabilities; High Risk Students; High Schools; *Instructional Effectiveness; *Language Arts; Poverty; Program Evaluation; *Reading Comprehension; Reading Difficulties; Reading Improvement; *Reading Instruction; Reading Strategies
IDENTIFIERS Considerate Instruction

ABSTRACT

This report discusses the findings of a study that evaluated the effectiveness of a set of considerate interventions in closing the language arts achievement gap in general, and evaluated the effects of these interventions in complex classroom settings that serve large numbers of at-risk students who have disabilities and live in poverty. The intervention was structured around the six instructional design principles of considerate instruction. The differences obtained in posttest performance of the at-risk groups (n=50) remaining in the study approached significance on the Multi-Level Academic Survey Test of reading comprehension favoring the considerate treatment. The students with disabilities (n=29) in the considerate at-risk classroom improved at a faster rate than their at-risk colleagues on the New Jersey Test of Reasoning. Overall, the results indicated that less than 50 hours of considerate instruction was not sufficient to narrow the achievement gap. The attempt to drop students with disabilities into the standards-based considerate instruction, regardless of prerequisite skills, failed. The two groups of at-risk students who began the considerate intervention in middle school and continued into high school, moving into the standards-based considerate programs with necessary prerequisite skills, seemed closer to attaining the desired outcome. (Contains 20 references, 10 tables, and 3 figures.) (Author/CR)



Institute for *Academic Access*

Research Report #13

**An experimental study of the effects of
considerate curricula in language arts on
reading comprehension and writing**

Bonnie Grossen, Shanna Hagen-Burke,
and Mack D. Burke

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to
improve reproduction quality.

• Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy.

2002



U.S. Office of Special
Education Programs

*Sponsored by the U.S. Department of Education,
Office of Special Education Programs
Grant # 84.324S*

BEST COPY AVAILABLE

Abstract

Impacting the low reading performance of high school students with disabilities is a monumental challenge. There is a growing literature that suggests the potential of "considerate" instruction for impacting the performance of these students. Considerate instruction is designed to incorporate the six principles of instructional design described by Kameenui and Carnine (2001). The two purposes of this study were to: (1) evaluate the effectiveness of a set of considerate interventions in closing the language arts achievement gap in general, and (2) evaluate the effects of these interventions in complex classroom settings that serve large numbers of at-risk students who have disabilities and live in poverty. The intervention was structured around the six instructional design principles of considerate instruction was delivered in large class settings. The differences obtained in posttest performance of the at-risk groups remaining in the study approached significance on the MAST test of reading comprehension (long form) favoring the considerate treatment. The SWDs in the considerate at-risk classroom improved at a faster rate than their at-risk colleagues on the New Jersey Test of Reasoning. Overall the results of this study indicated that with less than 50 hours of considerate instruction it is not sufficient to narrow the achievement gap. These findings underscore the fact that when SWDs are placed into classes without possessing the foundational pre-requisite skills, they struggle. The two groups of at-risk students who began the considerate intervention in middle school and continued into high school, moving into the standards-based considerate programs with the necessary prerequisite skills, seemed closer to attaining to reaching the desired outcome. This study points out that to be effective in closing the achievement gap for students who are significantly behind at the grade 9 level will require more time for considerate instruction and / or more intensity in the delivery of the instruction.

Closing the reading achievement gap between normally achieving students and students who fall behind in reading is especially challenging for older students. Hanushek, Kain, and Riukin (1998) found in an analysis of a large data set that pullout programs designed to close the achievement gap for special education students resulted in only a .04 standard deviation gain in reading performance. These insignificant gains are disappointing given the expense of creating individualized, specialized programs. However, they may not be surprising because studies of the instruction being provided in pullout programs (i.e., resource rooms) indicate that they are characterized by the same undifferentiated instruction that typically occurs in the mainstream (Moody, Vaughn, Hughes, & Fischer, 2000; Schumm & Moody, 2000). Nevertheless, even when teachers in resource settings used materials that supported differentiated instruction, there was no evidence of significant gains in language arts performance (Vaughn, Moody, & Shuman, 1998).

These and similar findings have been used to support the inclusion movement. Unfortunately, several studies have documented that inclusive practices are also ineffective in closing the achievement gap, especially for older students with reading deficits. For example, Klinger, Vaughn, Schumm, Hughes, and Elbaum (1997) found that 80% of the poorest readers made no measurable gain over an entire school year. Even when teachers received professional development and support, the amount of progress made by the end of the year did not narrow the reading gap (Foorman, Francis, Beeler, Winikates, & Fletcher, 1997).

A few researchers have shown that reading gains can be made by students if they receive intensive instruction. Torgesen et al. (2001), for example, documented one of the more powerful effects for an intervention for remedial readers. After 67.5 hours of instruction over an 8-week period, poor readers in grades 3 to 5 made a significant gain in reading achievement, which maintained for two years following the intervention. However, these students remained weak readers overall, showing practically no further improvement in their reading subsequent to the intervention. Although the students remained slow readers, the impact of intensive systematic instruction is noteworthy for its durability over time.

Turning poor reading performance around becomes even more difficult at the high-school level. Fuchs, Fuchs, and Kazdan (1999) examined the use of peer-assisted learning strategies (PALS) with high-school students with disabilities using a treatment-control group design. They found only modest growth in comprehension (an effect size of .34), little growth in reading fluency, and limited improvement in student attitudes and beliefs towards reading.

Fuchs et al. offer two specific reasons why high school students with reading-related disabilities have difficulty learning. Not only are the problems of older students more complicated, involving the emotional effects of years of failure, but high-school settings also pose a serious logistical challenge when scheduling special reading instruction. These logistical

challenges center on the problem that high schools generally do not provide any natural opportunities for reading instruction to occur in the mainstream.

Within this context, the effects found for "considerate" instruction on the learning of high school students with disabilities seem quite powerful (Grossen et al., 2002). "Considerate" instruction is instruction designed to incorporate the six principles of instructional design described by Kameenui and Carnine (2001). These six design principles accommodate the diverse learning needs of students with disabilities, children of poverty, and children with limited English while accelerating the learning of the group as a whole. Table 1 describes these six principles and contrasts them with traditional teacher-directed instruction.

Several studies found that considerate instruction had a significant effect on the reasoning and writing performance of high school students with disabilities. When a group of high school students with learning disabilities were taught using a considerate logic program, their performance matched that of high school students in an honors class and their ability to critique arguments exceeded the performance of college students enrolled in a teacher certification program (Collins & Carnine, 1988; Grossen & Carnine, 1990). In another study, the scores of mainstreamed high school students with learning disabilities did not differ significantly from control students without disabilities in the use of principles and facts in written analyses of primary source documents (Carnine, Caros, Crawford, Hollenbeck, & Harniss, 1996).

Similar to Torgesen et al.'s (2001) findings with younger children, these studies showed that intensive, systematic, considerate instruction can significantly change the performance of students with learning disabilities, even at the high school level, in the specific area of learning that was targeted by the instruction. However, these instructional interventions did not significantly change the overall performance of high school students with learning disabilities in the general domain of language arts.

Grossen and Carnine (1990) specifically investigated the problem of transfer of learning. They found that instruction can be designed so that the learning generalizes to untaught problems. However, as the problem types became less similar to the instructed problem types, performance diminished. This implies that to achieve an overall change in the performance of students with disabilities in the entire domain of language arts, intensive, systematic, considerate instruction would need to be provided in every significant skill area. Simply finding the instructional time for such a feat in the context of high school seems a daunting challenge. One purpose of the following study was to evaluate the effectiveness of a set of considerate interventions in closing the language arts achievement gap in general.

Furthermore, the previous studies evaluating considerate instruction represent level 2 research, where the variables can be better controlled (Ellis & Fouts, 1995, Grossen, 1996). For example, very well-trained, highly motivated teachers were used to deliver the instruction with

high fidelity. A second purpose of this study was to shift the evaluation of these interventions to level 3, in other words, to assess the effects of considerate instruction in the high schools that need them most—places where disruptive school contextual variables are more difficult to control. These high-need schools serve large numbers of at-risk students, not only students with disabilities but also those who live in poverty. Obtaining similar effect sizes in the context of high-need schools clearly presents its own set of challenges.

Considerate Instruction in High-Need Schools

Efforts to use considerate instruction to change the performance of diverse learners (students with disabilities, children of poverty, and children with limited English) in a large learning domain have focused primarily on the lower grades. A set of considerate instructional tools were identified by Carnine (1994) as “the BIG Accommodation Model.” The effects of the BIG Accommodation Model on the performance of diverse learners have been evaluated in some high-need middle schools in California (Grossen, 2002). With intensive effort, schools serving high-need populations have achieved significant gains in school-wide performance on standardized measures of language arts performance (Grossen, 2002, Grossen, in press). In addition, higher performing schools that implemented considerate instruction largely with their special education population, report significant gains in the performance of students with disabilities on standardized measures of general performance in the domain of language arts.

Research Questions

In the following study, we implemented language arts components of the BIG Accommodation Model of considerate instruction in various high-need high school settings to evaluate the effects on the performance of high frequency students with disabilities (SWDs). The SWDs in this study were students placed in a resource room for less than 50% of their school day. We also included groups of at-risk students in the analyses as a reference point for interpreting the performance of the SWDs. This was especially useful since our numbers of SWDs were often too low to allow the use of statistical tests of significant differences in the performance of groups.

Our analyses looked for answers to these questions: (1) Does the use of considerate instructional materials in language arts improve the performance of SWD's on measures of general language arts competence over teacher-prepared curricula? (2) Does the considerate instruction improve the performance of at-risk general education students on these measures as well?

Method

This research involved three high school settings. In the rural poverty setting, our methods were experimental. In the other two settings, urban poverty and suburban mixed, our methods were quasi-experimental. Because we were not able to carry out the experimental study

exactly as planned, the inclusion of our quasi-experimental data provides a stronger basis for making some inferences. We report the details regarding the methodology in two sections: experimental and quasi-experimental, after discussing the setting and measures, which are relevant to all the comparisons.

Settings

Three high school settings were used in the analyses: One high school in a rural area served a population with a large number of families living in poverty. Another suburban high school served a population of mixed socio-economic status (SES). A third large high school in the Midwest served a large urban population of families living in poverty. Table 2 displays critical demographic statistics for each of these high schools. Because other research indicates that intensity of instruction is an important variable of success, we include the scheduling format of the school as a possibly relevant factor in considering the effectiveness of instruction. In a block schedule students receive instruction every other day, with either a shortened period on Friday, or a regular period every other Friday. In a traditional schedule, students receive instruction every day, 5 days a week.

Measures

We used the following measures of general performance in the domain of language arts to evaluate learning:

SAT-9 statewide assessment data. California schools administer annually the Stanford Achievement Test, 9th edition, to evaluate student academic performance. The SAT-9 provides norm-referenced interpretations of performance in the areas of reading and language.

California High School Exit Exam (HSEE). The state of California required that a measure of academic proficiency designed by the state be taken by all 9th graders in the state. A pass score on this measure was designated as 350.

Reading Comprehension Benchmark. The *Multi-Level Academic Survey Test (MAST)* is a timed, norm-referenced, pencil-and-paper test that measures reading comprehension using a time-efficient maze procedure. The test provides norms for students in grade 2 through grade 12, providing a more sensitive measure of growth for older students who perform at the low end of the distribution than a traditional norm-referenced test. The MAST includes a short form, which requires 20 minutes to administer, and a long form, which requires approximately 40 minutes.

Because the measure includes easier items for the lower levels of reading proficiency a reliable distribution can be obtained at these lower levels for older lower-performing students. However, the grade level norms collapse the distribution into single digit percentiles for these lower performing older students. To provide a more sensitive measure of relative distribution, we used the performance of a national sample of grade 9 SWDs to calculate percentile scores for raw scores on the short form of the MAST. We used the short form because we had a more

complete data set for that form of the MAST. This sample included the complete population of grade 9 SWDs (total $n = 134$) at 9 high schools. SWDs are the higher performing students with disabilities who take at least half their coursework in mainstream classes. Although not selected with statistically rigorous consideration of representativeness in mind, the high schools nevertheless included a full range of the types of high schools found in America: 3 urban poverty schools, 3 rural poverty schools, and 3 suburban mixed schools. The schools also sampled a range of geographic areas: 3 were in Kansas, 2 in California, 1 in Oregon, and 3 in Washington state. The percentile scores referenced to the performance of this national sample of 9th grade SWDs provide an indicator of a student's relative rank in this sample.

Reasoning Benchmark. The *New Jersey Test of Reasoning Skills* (Shipman, 1983) is a 50-item pencil-and-paper test of elementary reasoning and inquiry skills. The reliability of the test for students in grades 7 and above is reported in the manual to be .91. The New Jersey Test was originally designed to assess the effectiveness of *Philosophy for Children*, a program for teaching critical thinking. An effort was made in creating the test "to construct a taxonomy in terms of the skills needed to perform the operations in the discipline of logic, both of the formal and informal variety, insofar as these relate to linguistic usage.... The taxonomy appears to be a reasonable representative of the domain, and the items selected for the New Jersey Test appear to be reasonably representative of the taxonomy" (Shipman, 1983, p. 14).

Integrity of Implementation Checklist. The research team designed this observation rubric to evaluate the integrity of the implementation of the considerate instruction on a scale of 0 to 3. All included groups reached at least the minimal criterion of a "2" on this scale. This is the minimal level of implementation fidelity that, based on anecdotal evidence, seems required to achieve significant results.

Experimental Comparisons

Subjects

A sample of 29 SWDs, who were placed for no more than 50% of the day in a special education class, were selected from a high school that served a rural population of high poverty families. To provide a point of reference and a context for evaluating the change in performance of the students with disabilities, we also included 50 at-risk students in the experimental study.

All of the students, both those with disabilities and those considered to be at-risk, had serious reading problems involving an inability to decode fluently or proficiently. Prior to treatment, these students all scored on the MAST reading subscale (short form) at a level lower than the average score achieved by the MAST grade 6 norming sample.

Procedures

Two special education teachers matched the students with disabilities in pairs based on the teachers' perceptions of their performance levels, and their classroom demeanor. After

independently ranking the students, the two teachers met together to resolve discrepancies. A student from each pair was randomly assigned to either the considerate treatment or the control group. The at-risk students were matched in pairs based on their performance on the curriculum-based *Corrective Reading Decoding Program* and assigned to treatment or control group. Five SWDs were taught with the at-risk students. The remaining 24 SWDs were taught in self-contained language arts classes designed specifically for SWDs—one an experimental class the other a control.

The experimental and control general education classes were taught by general education teachers and were designed specifically for students who were at risk of failure to graduate from high school. Both treatments were designed to teach to the standards of the high school exit exam; only the experimental treatment used the considerate instructional design. No more than one student with a disability was present in a general education class.

The experiment continued for one 9-week period on a block schedule. The contrasting experimental-and-control treatment was offered as one-half of a block, so the experimental and control groups received about 40 minutes of the contrasting instruction two and three times a week over a 9 week period. The total amount of instructional time was approximately 15 hours.

The Control Treatment

The control classes received instruction in "The Write Path," which emphasized connections to real life and is part of the AVID program for preparing at-risk students for college. Learning activities involved discussions about real-life problems and extensive opportunities to write about those problems using problem-solving frameworks.

The Experimental Treatment

The experimental group completed ten lessons of the considerate treatment, *Reasoning and Writing Level E* (Engelmann & Grossen, 1999), and the first mastery test. The teacher of the experimental groups was frequently ill and missed about 15 days of school, including days when coaching was offered to ensure that the teacher was delivering the instruction with fidelity. Consequently, the treatment group received approximately 10 hours of instruction from the trained teacher. However, for the ten lessons that were taught, the groups achieved the required mastery levels. (Eighty percent of the group is required to achieve at least 90% correct on the mastery test or the teacher reteaches specific segments of the lessons.) If the additional coaching had been provided as planned, the teacher might have become more efficient, achieving the required mastery levels with less reteaching. Nonetheless, because the group reached the mastery criteria on the first 10-lesson test, we can conclude that the program was implemented with the minimal required level of fidelity for the first 10 lessons.

Baseline Groups

A group of 25 SWDs representing the entire population of SWDs at the rural poverty school were given the MAST in April. Their scores are reported as a baseline comparison for the instructional groups: reading short form mean = 23.4 ($sd = 9.0$); reading long form mean = 34.2 ($sd = 13.6$). An additional group of at-risk students ($n = 37$) at the rural poverty school were also given the MAST. Their scores are used as a baseline for the at-risk students: reading short form mean = 31.8 ($sd = 6.7$), reading long form mean = 47.2 (10.4).

Results

The special education teacher of the self-contained class of SWDs implementing the experimental treatment dropped out of the study before completion. Only the pretest data were collected and those data are included in the data summaries in Table 3 below to see how their initial performance level may have differed from that of students who remained in the study. Their mean scores of the special education groups were significantly lower than those of the at-risk groups on the reading pretest, $t(37) = 2.04, p < .05$. The special education group seemed to lack the pre-skills necessary to learn from the standards-based program, although the SWDs who learned with the at-risk group scored lower initially than the experimental group of SWDs, and made stronger pre-to-post gains than the at-risk students with whom they learned.

Because the pretest scores were equivalent for the experimental and control at-risk general education groups, we used a simple analysis of variance to compare the posttest performance of the groups on reasoning and on reading. Neither difference was significant; however, the differences on the MAST reading comprehension measure (long form) approached significance, $t(1,48) = 1.32$.

Using the average standard deviations of the two treatment groups we calculated effect sizes as a measure of the amount of growth achieved by the SWDs in the respective treatments. The effect sizes on the New Jersey reasoning measure for the SWDs in the considerate treatment were consistently more than double that of the SWDs in the control treatment. On the reading comprehension measure (MAST long form), two of the three SWDs in the control group had negative effect sizes indicating a loss in overall reading competence. In one case the negative effect size was quite large ($-.82$). The one SWD with a positive effect size did not match the positive effect size of the considerate treatment group, nor did his effect size match that of the one SWD who took the reading posttest.

Table 4 displays these effect sizes. Figures 1 and 2 graphically display the slopes of the gains for the groups and for the individual students with disabilities. Visual inspection indicates a consistently steeper slope for the SWDs receiving the considerate instruction.

To evaluate the change in performance of the subjects of the study in terms of relative rank in the national sample of grade 9 SWDs, we converted the MAST short form raw scores to

the percentile scores derived from that national SWD sample. Table 5 displays these scores. Only the scores for the experimental considerate treatment show a gain. The scores for the control group and the SWDs in the control treatment were all negative or zero.

Quasi-Experimental Comparisons

Subjects

Six grade 9 SWDs from the suburban mixed SES public high school participated in the quasi-experimental study. Three SWDs, who were assigned to a special education placement for less than 50% of their instruction, attended the general education language arts classes for at-risk students that were taught using the considerate instructional design. Three attended special education language arts classes designed specifically for special education students.

The at-risk grade 9 students who received considerate instruction and the grade 9 normally achieving general education students were also included in the study. In addition, quasi-experimental comparisons were made with other instructional groups from the other high school settings.

Procedures

At-risk students were placed into one of three types of considerate instruction based on their performance. The treatment groups were as follows:

Considerate Remedial (ConsRem).

Students who received remedial instruction were low readers who began receiving considerate instruction in grade 9 (*Corrective Reading—Decoding*, Engelmann and associates, 1999). These students failed to meet minimum criteria on a decoding test (approximately grade 6 level).

Considerate Standards-based 1 (ConsSB1).

Students who received standards-based instruction were low readers who began receiving considerate instruction in middle school and continued into high school to receive standards-based language arts instruction. The specific instructional program was the *Corrective Reading—Comprehension Level C* program (Engelmann and associates, 1999).

Considerate Standards-based 2 (ConsSB2).

As in the above group, these students began receiving considerate instruction in middle school. However, the specific instructional program these students received was *Reasoning and Writing Level E* (Engelmann & Grossen, 2001). This is the same program used in the considerate treatment in the experimental study. These students had already taken the first standards-based course and, in contrast to the experimental groups, entered the course with a higher pre-skill level of performance.

General Education Regular Curriculum (ControlReg).

The performance of normally achieving students receiving instruction in the regular high school language arts curriculum was compared with the performance of the groups receiving considerate instruction.

Of the three SWDs placed into the considerate treatments, two of them were placed with the at-risk group receiving considerate standards-based instruction. One received instruction with the at-risk group receiving considerate remedial reading instruction. The three SWDs receiving the control treatment received their instruction in special education pullout classes designed specifically for their needs. They were not placed in the regular education control group. All of the students in the above groups were from the suburban mixed high school. All groups with pre-and-posttest scores received approximately 25 hours of instruction between those testing occasions.

Other Comparison Groups

We also examined the following additional groups of at-risk students receiving considerate instruction in other settings.

1. *Traditional schedule considerate remedial instruction (ConsDailyRem).* A sample of 26 at-risk students from a high poverty urban high school in a Midwestern state, who received the considerate treatment in a traditional schedule of daily instruction for 45 minutes a day, was selected for comparison. All students in grades 9 to 11 who scored on the MAST reading subscale (short form) at a level lower than the average score achieved by the MAST norming sample of grade 6 students were selected for the sample. These 26 students included 17 grade 9 students, 7 grade 10 students, and 2 grade 11 students.

The 26 subjects received considerate instruction in basic reading skills during one 45-minute period per day for 5 months—approximately 50 hours of instruction. The considerate program was *Corrective Reading—Decoding*. Students were grouped into three different levels with students at a similar reading level for instruction. The MAST was administered as a pretest in November and a posttest in April.

2. *Block schedule considerate standards-based instruction (ExpAt-Risk).* The performance of the group receiving considerate instruction in the experimental comparison is included here also for quasi-experimental comparisons. This group received 10 hours of instruction every other day for a 9-week quarter and covered 10 lessons of the standards-based considerate instruction with mastery.

3. *Block schedule control standards-based instruction (ControlAt-Risk).* The performance of the group receiving the control treatment (the AVID program) in the experimental study is also included here for quasi-experimental comparisons. This group received about 15 hours of instruction on a block schedule every other day for 9 weeks.

In addition, the mean scores of the baseline SWD and at-risk groups for the rural poverty school are included in the comparisons.

Results

Group Differences on the SAT-9 and HSEE

To evaluate the effects of the treatments on general language arts outcomes, we collected raw scores on the SAT-9, calculated the means for the various groups, and then converted the mean raw score to a percentile. We used only matched scores so students without a score from the previous year were not included in the analysis. Students who received the considerate remedial instruction were generally new to the district, and did not have SAT-9 scores for the previous year, so the considerate remedial group is not included in this analysis. Similarly, many of the subjects in the control regular education class were also new to the district. Generally, only the students returning to the district who have never been in need of remedial instruction were in the control regular education group.

Table 6 displays the results of these analyses. There was a sharp drop in the reading percentile equivalents from the previous year for all groups. The control regular education group experienced the steepest decline, a loss of 13 percentile points. The considerate instructional groups declined 8 and 10 percentile points.

We checked the consistency of the percentile scores with a larger sample, over 300,000 grade 9 students in California. The percentile for the mean raw score for all grade 9 students tested in reading declined 12 points, while the language score remained the same. This large drop for such a large population raises questions about the representativeness of the norming sample for the SAT-9. With such a large sample the percentile score for the mean raw score should remain at about the 50th percentile from one year to the next, if the norming sample was representative.

The change in language percentile scores varied across the instructional groups of our study. The scores of the considerate standards-based1 group (the initially lower-performing standards-based group) showed an increase of 10 percentile points; the considerate standards-based2 group remained flat; and the scores of the control regular education group dropped 1 percentile point.

To test for differences among the groups in learning gains as measured by the SAT-9 we used an analysis of covariance (ANCOVA) with the mean raw scores from the previous year as the concomitant variable. Table 6 displays the means and standard deviations for matched scores for the groups represented in this analysis. According to the ANCOVA, the groups did not differ significantly in slope or elevation.

In addition, the California High School Exit Exam was administered to all grade 9 students. Table 7 displays the means and standard deviations of the scores for the three groups.

An additional MANCOVA test, using the SAT-9 reading raw score as the concomitant variable, indicated significant differences among groups. A Scheffe' test indicated that the only significant difference was between the considerate standards-based1 treatment and the control regular education treatment on the California High School Exit Exam, favoring the control regular education treatment.

The mean score on the HSEE for all groups was higher than the pass score of 350.

The Performance of Students With Disabilities on the SAT-9 And HSEE

We evaluated the change in performance of the students with disabilities by charting their percentile scores on the SAT-9 reading and language subscales. Table 8 presents these scores. SWDs 1 and 2 were taught in the considerate standards-based 1 treatment, SWD 3 was taught in the considerate remedial treatment, and the control SWDs were taught in the resource room. Figure 3 graphically displays the change in performance of these students as indicated by the percentile scores on the language subscale of the SAT-9. Of the 12 pre-to-post scores for SWDs, only two scores showed a rise from pretest to posttest. Those were for two SWDs in the area of language. Only one SWD received a pass score. That student (SWD 11) was in the control group and also achieved the highest score on the pretest. SWD 7 nearly passed. SWD 7 had been in the considerate treatment for two years prior at the middle school level. By grade 9 he nearly passed the HSEE on the first try.

Performance on the MAST

Percentile equivalents using grade level norms, and grade level equivalents are also reported. Because the pre-and-posttests included a shorter interval than a full year, comparing the pretest to grade 8 norms and the posttest to grade 9 norms, is somewhat unfair. The change in raw scores is displayed to show that some growth occurred. The only group showing a rise in percentile score was the considerate standards-based2 group. We used a MANCOVA to test for differences among the groups, using the pretest as the concomitant variable. The results indicated that the considerate standards-based2 group scored significantly higher on the posttest than the control group ($p < .05$). No other differences were significant.

To determine the relative rank of the various groups and the individual SWDs in the national sample of grade 9 SWDs we used the MAST short form scores. Table 10 displays each SWD and each group's relative rank with the national sample of grade 9 SWDs as a percentile. The group receiving considerate daily remedial instruction (ConsDailyRem) made significantly greater gains than.

Discussion

In this study we faced two major new challenges. We tested whether we could narrow the achievement gap experienced by high school SWDs and at-risk students in the broad domain of language arts. Our measures, therefore, covered a broad domain, rather than a narrow domain as

in previous short duration studies. Second, we tested the most powerful intervention model we know in high-need schools. Consequently, the teachers were not experts in the model working in normally achieving schools, as in our previous studies, but the teachers working and living in that high-need environment every day.

The Rural Poverty Setting

In the rural poverty setting, the experimental comparison of a relatively large number of SWDs was dropped before completion. SWDs were placed in the self-contained class versus the general education class based on their IEP, which could mean that the self-contained students were lower performing. To determine whether the preskills of the self-contained group of SWDs may have been lower, thus making the standards-based intervention too difficult for them, we compared their pretest scores to the rest of the groups. The performance level of these SWDs was lower than that of the other groups. Further research is needed to explain why two SWDs learning in the general education class seemed to do quite well in spite of having less initial reading competence than the group of SWDs who could not manage the material.

The differences in posttest performance of the at-risk groups remaining in the study approached significance on the MAST test of reading comprehension (long form) favoring the considerate treatment. Interestingly, the SWDs in the considerate at-risk classroom seemed to improve at a faster rate than their at-risk colleagues on the New Jersey Test of Reasoning. Unfortunately, only one SWD was present for the MAST reading posttest.

Possibly a longer duration for the study would have produced significant results. We attempted to continue the study into the second semester; however, the circumstances of the high school schedule required extensive changes in each student's program. It was impossible within the current high school framework to extend the intervention.

The Suburban Mixed SES Setting

In contrast to the experimental study in the rural poverty school, where special education students were randomly assigned to the standards-based instruction (*Reasoning and Writing*), students from the suburban setting were placed in instructional groups based on their performance on a placement test. The students placed in the standards-based 2 (ConsSB2) program, *Reasoning and Writing*, were only able to do so because they had taken considerate courses at the middle school level that prepared them for this program. Consequently, the students in the higher group (SB2) were able to achieve scores that were well within passing range of HSEE. Although the mean score of the ConsSB2 group was nonsignificantly lower than the control regular education group, the standard deviation was smaller. The score one standard deviation below the mean for the SB2 group (358) was higher than that of the control group (351). More students in the SB2 group passed the HSEE than in the control group: 52% of the SB1 group passed; 88% passed in the SB2 group; and 81% passed in the control group.

The higher level standards-based program unfortunately contained no SWDs. Two SWDs were included in the SB1 group, and one of those came within 2 points of passing the HSEE. Similarly, one SWD in the control group passed the HSEE by 4 points.

Conclusions

Overall the results of our study indicated that with less than 50 hours of considerate instruction, we can barely begin to narrow the achievement gap when the students begin the solution at grade 9 level. The attempt to drop SWDs into the standards-based considerate instruction, regardless of whether they had the prerequisite skills, failed. Our two groups of at-risk students who began the considerate intervention in middle school and continued into high school, moving into the standards-based considerate programs with the necessary prerequisite skills, seemed closer to attaining our goal.

To begin to close the achievement gap for students who are significantly behind at the grade 9 level will require more time for considerate instruction and / or more intensity in the delivery of the instruction. We noted anecdotally that there was room to increase the intensity of the instruction in most of the classrooms. Many minutes of precious instructional time were wasted, especially in high schools with block schedules.

In addition, high school models that allow more instructional time for accelerated catch up using considerate programs are needed. The rural poverty high school actually has begun a program where students who are significantly behind are pulled out of the regular high school setting and placed in an alternative high school, where a full-day curriculum of considerate instruction is offered until the students reach critical benchmarks. When they reach these benchmarks, they may re-enter the regular high school. Students with disabilities, however, are placed in the regular high school. They, of course, already have a pullout program.

In summarizing what we know about closing the learning gap for students with disabilities, Lyon et al. (2001) suggested that remediation models for older children have been ineffective for two main reasons. "First, the instruction provided through remediation is frequently too little, too general, and too unsystematic. Secondly, even if the instruction were of high quality, it may be too late given that many children are already far behind and less motivated to learn to read" (p. 272). An intensive schedule with considerate instructional programs seems to have the most promise for closing the gap for high school SWDs. We found no shortcuts.

References

- Carnine, D. (March, 1994). The BIG Accommodation program. *Educational Leadership*, 87-88.
- Carnine, D., Caros, J., Crawford, D., Hollenbeck, K., Harniss, M. (1996). Designing effective U.S. history curricula for all students. In J. Brophy (Ed.), *6th Handbook on Research on Teaching*. Advances in Research on Teaching (Vol.6), 207-256. Greenwich, CT: J.A.I. Press.
- Caros, J. (2001). *Emerald Ice 4th-dimension database*. Vancouver, BC: Seira.
- Collins, M., & Carnine, D. (1988). Evaluating the field test revision process by comparing two versions of a reasoning skills CAI program. *Journal of Learning Disabilities*, 21, 375-379.
- Engelmann, S., & Grossen, B. (1999). *Reasoning and writing level E*. Blacklick, OH: Science Research Associates.
- Foorman, B., Francis, D., Beeler, T., Winikates, D., & Fletcher, J. (1997). Early interventions for children with reading problems: Study designs and preliminary findings. *Learning Disabilities*, 8, 63-71.
- Fuches, L. S., Fuches, D., and Kazdan, S. (1999). Effects of peer-assisted learning strategies on high school students with serious reading problems. *Remedial and Special Education*, 20(5), 309-319.
- Grossen, B. (1998). The research base for *Corrective Reading*, SRA. Blacklick, OH: SRA.
- Grossen, B. (2002). The BIG Accommodation Model. *Journal for the Education of Students Placed At Risk*, 7(2), 241-263.
- Grossen, B. (in press). Success of a direction instruction model at a secondary level school with high-risk students. *Reading and Writing Quarterly*.
- Grossen, B., & Carnine, D. (1990). Diagramming a logic strategy: Effects on more difficult problem types and transfer. *Learning Disability Quarterly*, 13, 168-182.
- Grossen, B., Caros, J., Carnine, D., Davis, B., Deshler, D., Schumaker, J., Bulgren, J., Lenz, K., Adams, G., Jantzen, J., & Marquis, J. (2002). Big ideas (plus a little effort) produce big results. *Teaching Exceptional Children*, 34(4), 70-73.
- Grossen, B., Lee, C., & Johnston, D. (1995). A comparison of the effects of direct instruction in reasoning with constructivism on deductive reasoning. In A. Deffenbaugh, G. Sugai, & G. Tindal (Eds.), *The Oregon Conference 1995 Monograph* (pp. 253-273). Eugene, OR: College of Education, University of Oregon.
- Hanushek, E., Kain, J., & Riukin, S. (1998). *Does special education raise academic achievement for students with disabilities?* National Bureau of Economic Research working paper no. 6469.

- Kameenui, E., & Carnine, D. (2001). *Effective teaching strategies that accommodate diverse learners*. Columbus, OH: Merrill.
- Lyon, R., Fletcher, J., Shaywitz, S., Shaywitz, B., Torgesen, J., Wood, F., Schulte, A., & Olson, R. (2001). *Rethinking learning disabilities*. Progressive Policy Institute, Fordham Foundation.
- Moody, s., Vaughn, S., Hughes, M. & Fischer, M. (2000). Reading instruction in the resource room: Set up for failure. *Exceptional Children*, 16, 305-316.
- Schumm, J., Moody, S., & Vaughn, S. (2000). Grouping for reading instruction: Does one size fit all? *Journal of Learning Disabilities*, 5_477-488.
- Torgesen, J., Alexander, A., Wagner, R., Rashotte, C., Voeller, K., Conway, T., & Rose, E. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, 34, 33-58.
- Vaughn, S., Moody, S., & Schumm, J. (1998). Broken promises: Reading instruction in the resource room. *Exceptional Children*, 64, 211-225.

Table 1.

The Contrast Between Instruction with Accommodations for Diverse Learners and Traditional Instruction

Six Principles of Accommodation for Diverse Learners	Traditional Instruction
<u>Big Ideas</u> , concepts and principles that facilitate the most efficient and broad acquisition of knowledge across a range of examples, are presented. Big ideas make it possible for students to learn the most and learn it as efficiently as possible, because "small" ideas can often be best understood in relationship to larger, "umbrella concepts."	A barrage of unrelated facts and details are presented. The links between concepts are obscured.
<u>Conspicuous Strategies</u> made up of specific steps that lead to solving complex problems are taught.	Strategies are seldom taught.
<u>Background Knowledge</u> is pre-taught.	Important prerequisite learning is often not evaluated nor taught.
<u>Mediated Scaffolding</u> provides personal guidance, assistance, and support that gradually fades as students become more proficient and independent.	Little direction or provision for scaffolding the progression of learning toward greater independence is provided.
<u>Judicious Review</u> requires students to draw upon and apply previously learned knowledge over time.	Review is often minimal.
<u>Strategic Integration</u> blends new knowledge with old knowledge to build bigger big ideas.	Spiraling of topics does not carefully integrate units.

Table 2
Demographic Characteristics for Three High School Settings

Characteristic	Rural Poverty	Suburban Mixed	Urban Poverty
Participants eligible for free or reduced lunch	57%	24%	44%
African-Americans	5%	28%	17%
Latinos	53%	32%	12%
Asian	1%	14%	14%
Total school enrollment	498	635	2181
Location	California	California	Midwest
Scheduling format	Block (groups meet every other day)	Block (groups meet every other day)	Traditional (groups meet every day)

Table 3.
Descriptive Statistics of Performance of Students with Disabilities (SWD) and the Comparison Groups on the New Jersey Test of Reasoning and on the Multi-Level Academic Survey Test

Group	n	Pre-Test		Post-Test		Rank in Group
		Reasoning ¹	Reading ²	Reasoning	Reading	
Considerate Language Arts Instruction						
SWD 1		12	32	20	41	19->8
SWD 2		14	--	23	--	13->4
Exp Sped	11	13.4 (3.8)	36.5 (10.0)			
Exp At-Risk	28	13.9 (4.7)	46.3 (10.7)	16.0(4.7)	49 (8.3)	
Control Language Arts Instruction						
SWD 3		8	32	10	29	21->19
SWD 4		11	35	14	27	13->14
SWD 5		14	48	17	50	9->4
Ctrl Sped	13	13.6 (3.7)	36.0 (11.1)	--	--	
Ctrl At-Risk	22	13.7 (4.7)	46.6 (8.9)	15 (4.9)	45.4 (11.0)	
Baseline SWD	25				34.2 (12.6)	
Baseline At-Risk	37				47.2 (10.4)	

¹Reasoning= New Jersey Test of Reasoning (Shipman, 1983)

²Reading= MAST

Table 4

Effect Sizes for Pre- to Post-test Achievement Gain for Students with Disabilities and the Comparison Groups on the New Jersey Test of Reasoning and on the MAST Reading Test (long form)

Students	Effect Size	
	Reasoning (sd = 4.7)	Reading (sd = 9.8)
Experimental at-risk group	.45	.28
Exp SWD 1	1.7	.92
Exp SWD 2	1.9	--
Control at-risk group	.28	-.12
Control SWD 3	.43	-.31
Control SWD 4	.64	-.82
Control SWD 5	.64	.20

Note: Standard deviations for the denominator were calculated by averaging the standard deviations for the two groups on the pretest.

Table 5.

Performance on the MAST Short Form Reported in Mean Raw Scores with Standard Deviations, and Percentile Scores Derived from a National Sample of SWDs.

	Pretest		Posttest		Change in %tile Rank
	Mean (sd)	%tile	Mean(sd)	%tile	
Exp					
SWD 1	22	26	29	41	15
Exp SWD Group (n = 11)	26.8 (8.2)	36	--	--	--
Exp At-Risk Group (n=28)	31.5 (5.7)	49	33.6 (3.8)	54	5
Control					
Control SWD 3	28	38	23	28	- 10
Control SWD 4	27	36	21	25	- 2
Control SWD 5	36	65	36	65	0
Control SWD Group (n = 13)	25.3 (6.0)	33	--	--	--
Control At-Risk (n=22)	32.8 (4.2)	52	31.4 (6.4)	49	- 3
Baseline SWD (n=25)	--	--	23.4 (9.0)	29	--
Baseline At-Risk (n=37)	--	--	31.8 (6.7)	49	--

Table 6

Percentiles for the Mean Raw Matched Scores on the Reading and Language Subscales of the SAT-9 for Instructional Groups in the Suburban Mixed SES High School

Treatment	N	Percentile of the SAT-9 Mean Raw Matched Score			
		Reading		Language	
		Pre	Post	Pre	Post
ConsSB1	59	35	25	31	41
ConsSB2	23	51	43	60	60
Control Reg	103	54	41	58	57

Table 7

Mean Raw Scores and Standard Deviations for the Performance of the Treatment Groups on the SAT-9

		Mean Raw Scores (SD)				
Group	N	Reading		Language		HSEE
		Pre	Post	Pre	Post	Post
ConsSB1	59	51 (11)	45 (14)	27 (7)	25 (9)	355 (29)
ConsSB2	23	58 (8)	56 (9)	34 (4)	31 (5)	376 (18)
Control Reg	103	60 (12)	55 (14)	33 (8)	30 (10)	383 (32)

Table 8
Percentile Scores of Students with Disabilities on the SAT-9 Reading and Language Arts

Group	SAT-9 Reading		SAT-9 Language		HSEE
	Pre	Post	Pre	Post	Post
Experimental					
Cons SWD 6	13	6	20	38	305
Cons SWD 7	14	14	41	28	348
Cons SWD 8	19	4	11	14	317
Control					
Ctrl SWD 9	9	3	14	11	307
Ctrl SWD 10	10	4	5	1	300
Ctrl SWD 11	26	14	29	23	354

Table 9

Performance of the Treatment Groups on the Multi-Level Academic Survey Test (Long Form) in Mean Raw Score, Standard Deviation, Percentile Equivalent, and Grade Equivalent.

Group	N	Pretest		Posttest		Raw Score
		Mean (SD)	PR ¹ / GE	Mean (SD)	PR ² / GE	Change
Cons SWD 6		35	5 / 3.1	30	2 / 2.7	- 5
Cons SWD 7		42	6 / 4.7	43	7 / 4.9	+ 1
Cons SWD 8		52	28 / 6.7	53	24 / 7.1	+ 1
Cons SB1	28	54.3 (6.3)	34 / 7.3	55.1 (10.4)	32 / 7.4	+ 0.8
Cons SB2*	19	56.7 (7.6)	44 / 7.8	59.1 (6.6)	50 / 9.0	+ 2.4
Control At-Risk*	22	46.6 (8.9)	13 / 5.5	45.4 (11.0)	9 / 5.1	-1.2
Control Rural SWD	25	--	--	34.2 (12.6)	3 / 2.9	--

¹ Percentile value using end of grade 8 norms

² Percentile value using end of grade 9 norms

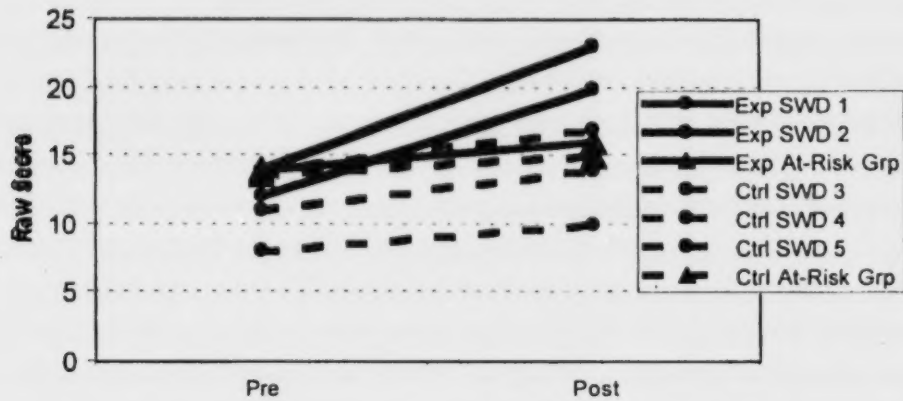
*Difference between these two groups significant at a $p < .05$ level.

Note: Exp SWD 6 received the ConsRem treatment. Exp SWDs 7 and 8 received the ConsSB1 treatment.

Table 10.

Descriptive Statistics of the Performance of Quasi-Experimental Comparison Groups on the MAST (Short Form) and of Individual SWDs with Percentile Scores Derived from a National Sample.

	N	Pretest Mean (SD)	SWD %ile	Posttest Mean (SD)	SWD %ile	%ile Change
<u>Experimental on Daily Schedule</u>						
ConsDailyRem	26	24.4 (8.9)	31	29.3 (9.5)	42	+ 11
<u>Considerate on Block Schedule</u>						
Cons SWD 6	1	21	25	24	30	+ 5
Cons SWD 7	1	34	55	33	53	- 2
Cons SWD 8	1	32	51	37	69	+ 18
ConsBlockSB1	23	33.6 (6.1)	53	36.0 (4.5)	63	+ 10
ConsBlockSB2	19	36.8 (3.5)	68	38.2 (1.9)	75	+ 7
ExpAtRisk	28	31.5 (5.7)	49	33.6 (3.8)	54	+ 5
<u>Control Groups on Block Schedule</u>						
ControlAtRisk	22	32.8 (4.2)	52	31.4 (6.4)	49	- 3
BaselineRuralSWDs	25	--	--	23.4 (9.0)	30	--
BaselineRuralAtRisk	37	--	--	31.8 (6.7)	49	--



Note: Groups are indicated with a triangular-shaped point. Individual students with disabilities are indicated with a circular point.

Figure 1. Graphic display of the pre-to-post gain for comparison groups and for students with disabilities on the New Jersey Test of Reasoning.

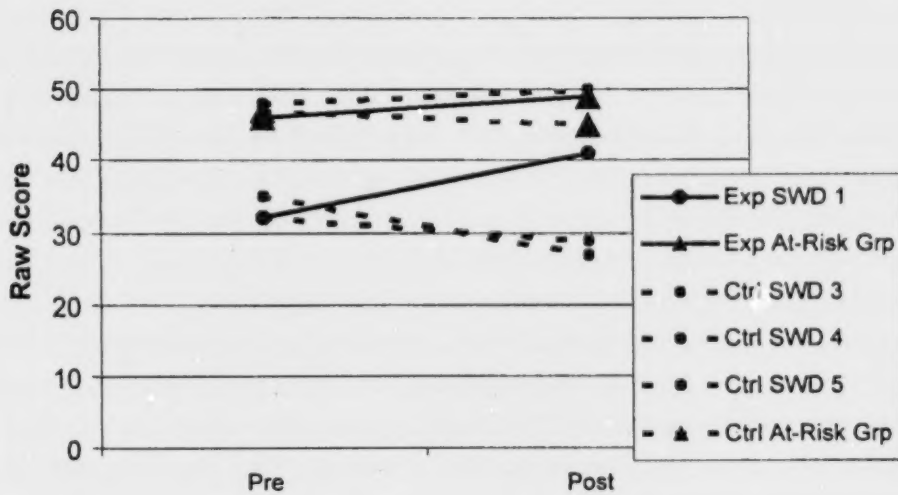


Figure 2. Graphic display of pre-to-post gain for comparison groups and for students with disabilities on the Multi-Level Academic Survey Test.

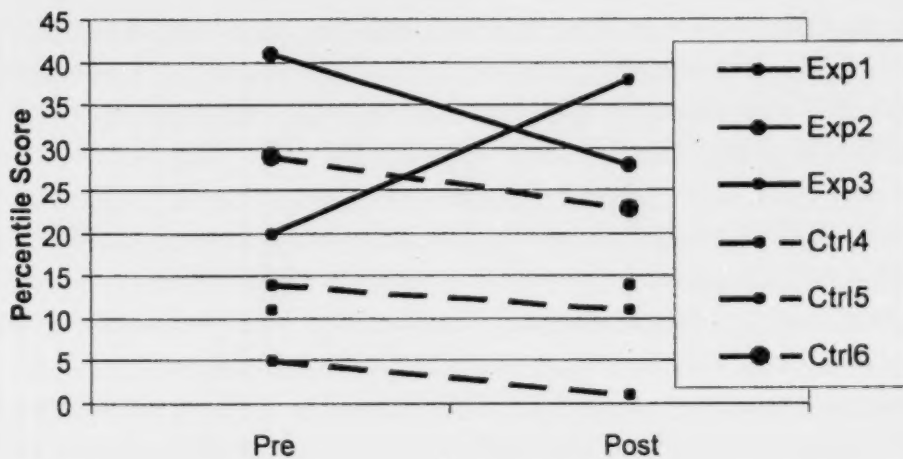


Figure 3. Graphic display of change in performance of students with disabilities on standards-based measures (SAT-9).

Abstract

Impacting the low reading performance of high school students with disabilities is a monumental challenge. There is a growing literature that suggests the potential of "considerate" instruction for impacting the performance of these students. Considerate instruction is designed to incorporate the six principles of instructional design described by Kameenui and Carnine (2001). The two purposes of this study were to: (1) evaluate the effectiveness of a set of considerate interventions in closing the language arts achievement gap in general, and (2) evaluate the effects of these interventions in complex classroom settings that serve large numbers of at-risk students who have disabilities and live in poverty. The intervention was structured around the six instructional design principles of considerate instruction was delivered in large class settings. The differences obtained in posttest performance of the at-risk groups remaining in the study approached significance on the MAST test of reading comprehension (long form) favoring the considerate treatment. The SWDs in the considerate at-risk classroom improved at a faster rate than their at-risk colleagues on the New Jersey Test of Reasoning. Overall the results of this study indicated that with less than 50 hours of considerate instruction it is not sufficient to narrow the achievement gap. These findings underscore the fact that when SWDs are placed into classes without possessing the foundational pre-requisite skills, they struggle. The two groups of at-risk students who began the considerate intervention in middle school and continued into high school, moving into the standards-based considerate programs with the necessary prerequisite skills, seemed closer to attaining to reaching the desired outcome. This study points out that to be effective in closing the achievement gap for students who are significantly behind at the grade 9 level will require more time for considerate instruction and / or more intensity in the delivery of the instruction.

Closing the reading achievement gap between normally achieving students and students who fall behind in reading is especially challenging for older students. Hanushek, Kain, and Riukin (1998) found in an analysis of a large data set that pullout programs designed to close the achievement gap for special education students resulted in only a .04 standard deviation gain in reading performance. These insignificant gains are disappointing given the expense of creating individualized, specialized programs. However, they may not be surprising because studies of the instruction being provided in pullout programs (i.e., resource rooms) indicate that they are characterized by the same undifferentiated instruction that typically occurs in the mainstream (Moody, Vaughn, Hughes, & Fischer, 2000; Schumm & Moody, 2000). Nevertheless, even when teachers in resource settings used materials that supported differentiated instruction, there was no evidence of significant gains in language arts performance (Vaughn, Moody, & Shuman, 1998).

These and similar findings have been used to support the inclusion movement. Unfortunately, several studies have documented that inclusive practices are also ineffective in closing the achievement gap, especially for older students with reading deficits. For example, Klinger, Vaughn, Schumm, Hughes, and Elbaum (1997) found that 80% of the poorest readers made no measurable gain over an entire school year. Even when teachers received professional development and support, the amount of progress made by the end of the year did not narrow the reading gap (Foorman, Francis, Beeler, Winikates, & Fletcher, 1997).

A few researchers have shown that reading gains can be made by students if they receive intensive instruction. Torgesen et al. (2001), for example, documented one of the more powerful effects for an intervention for remedial readers. After 67.5 hours of instruction over an 8-week period, poor readers in grades 3 to 5 made a significant gain in reading achievement, which maintained for two years following the intervention. However, these students remained weak readers overall, showing practically no further improvement in their reading subsequent to the intervention. Although the students remained slow readers, the impact of intensive systematic instruction is noteworthy for its durability over time.

Turning poor reading performance around becomes even more difficult at the high-school level. Fuchs, Fuchs, and Kazdan (1999) examined the use of peer-assisted learning strategies (PALS) with high-school students with disabilities using a treatment-control group design. They found only modest growth in comprehension (an effect size of .34), little growth in reading fluency, and limited improvement in student attitudes and beliefs towards reading.

Fuchs et al. offer two specific reasons why high school students with reading-related disabilities have difficulty learning. Not only are the problems of older students more complicated, involving the emotional effects of years of failure, but high-school settings also pose a serious logistical challenge when scheduling special reading instruction. These logistical

challenges center on the problem that high schools generally do not provide any natural opportunities for reading instruction to occur in the mainstream.

Within this context, the effects found for "considerate" instruction on the learning of high school students with disabilities seem quite powerful (Grossen et al., 2002). "Considerate" instruction is instruction designed to incorporate the six principles of instructional design described by Kameenui and Carnine (2001). These six design principles accommodate the diverse learning needs of students with disabilities, children of poverty, and children with limited English while accelerating the learning of the group as a whole. Table 1 describes these six principles and contrasts them with traditional teacher-directed instruction.

Several studies found that considerate instruction had a significant effect on the reasoning and writing performance of high school students with disabilities. When a group of high school students with learning disabilities were taught using a considerate logic program, their performance matched that of high school students in an honors class and their ability to critique arguments exceeded the performance of college students enrolled in a teacher certification program (Collins & Carnine, 1988; Grossen & Carnine, 1990). In another study, the scores of mainstreamed high school students with learning disabilities did not differ significantly from control students without disabilities in the use of principles and facts in written analyses of primary source documents (Carnine, Caros, Crawford, Hollenbeck, & Harniss, 1996).

Similar to Torgesen et al.'s (2001) findings with younger children, these studies showed that intensive, systematic, considerate instruction can significantly change the performance of students with learning disabilities, even at the high school level, in the specific area of learning that was targeted by the instruction. However, these instructional interventions did not significantly change the overall performance of high school students with learning disabilities in the general domain of language arts.

Grossen and Carnine (1990) specifically investigated the problem of transfer of learning. They found that instruction can be designed so that the learning generalizes to untaught problems. However, as the problem types became less similar to the instructed problem types, performance diminished. This implies that to achieve an overall change in the performance of students with disabilities in the entire domain of language arts, intensive, systematic, considerate instruction would need to be provided in every significant skill area. Simply finding the instructional time for such a feat in the context of high school seems a daunting challenge. One purpose of the following study was to evaluate the effectiveness of a set of considerate interventions in closing the language arts achievement gap in general.

Furthermore, the previous studies evaluating considerate instruction represent level 2 research, where the variables can be better controlled (Ellis & Fouts, 1995, Grossen, 1996). For example, very well-trained, highly motivated teachers were used to deliver the instruction with

high fidelity. A second purpose of this study was to shift the evaluation of these interventions to level 3, in other words, to assess the effects of considerate instruction in the high schools that need them most—places where disruptive school contextual variables are more difficult to control. These high-need schools serve large numbers of at-risk students, not only students with disabilities but also those who live in poverty. Obtaining similar effect sizes in the context of high-need schools clearly presents its own set of challenges.

Considerate Instruction in High-Need Schools

Efforts to use considerate instruction to change the performance of diverse learners (students with disabilities, children of poverty, and children with limited English) in a large learning domain have focused primarily on the lower grades. A set of considerate instructional tools were identified by Carnine (1994) as “the BIG Accommodation Model.” The effects of the BIG Accommodation Model on the performance of diverse learners have been evaluated in some high-need middle schools in California (Grossen, 2002). With intensive effort, schools serving high-need populations have achieved significant gains in school-wide performance on standardized measures of language arts performance (Grossen, 2002, Grossen, in press). In addition, higher performing schools that implemented considerate instruction largely with their special education population, report significant gains in the performance of students with disabilities on standardized measures of general performance in the domain of language arts.

Research Questions

In the following study, we implemented language arts components of the BIG Accommodation Model of considerate instruction in various high-need high school settings to evaluate the effects on the performance of high frequency students with disabilities (SWDs). The SWDs in this study were students placed in a resource room for less than 50% of their school day. We also included groups of at-risk students in the analyses as a reference point for interpreting the performance of the SWDs. This was especially useful since our numbers of SWDs were often too low to allow the use of statistical tests of significant differences in the performance of groups.

Our analyses looked for answers to these questions: (1) Does the use of considerate instructional materials in language arts improve the performance of SWD’s on measures of general language arts competence over teacher-prepared curricula? (2) Does the considerate instruction improve the performance of at-risk general education students on these measures as well?

Method

This research involved three high school settings. In the rural poverty setting, our methods were experimental. In the other two settings, urban poverty and suburban mixed, our methods were quasi-experimental. Because we were not able to carry out the experimental study

exactly as planned, the inclusion of our quasi-experimental data provides a stronger basis for making some inferences. We report the details regarding the methodology in two sections: experimental and quasi-experimental, after discussing the setting and measures, which are relevant to all the comparisons.

Settings

Three high school settings were used in the analyses: One high school in a rural area served a population with a large number of families living in poverty. Another suburban high school served a population of mixed socio-economic status (SES). A third large high school in the Midwest served a large urban population of families living in poverty. Table 2 displays critical demographic statistics for each of these high schools. Because other research indicates that intensity of instruction is an important variable of success, we include the scheduling format of the school as a possibly relevant factor in considering the effectiveness of instruction. In a block schedule students receive instruction every other day, with either a shortened period on Friday, or a regular period every other Friday. In a traditional schedule, students receive instruction every day, 5 days a week.

Measures

We used the following measures of general performance in the domain of language arts to evaluate learning:

SAT-9 statewide assessment data. California schools administer annually the Stanford Achievement Test, 9th edition, to evaluate student academic performance. The SAT-9 provides norm-referenced interpretations of performance in the areas of reading and language.

California High School Exit Exam (HSEE). The state of California required that a measure of academic proficiency designed by the state be taken by all 9th graders in the state. A pass score on this measure was designated as 350.

Reading Comprehension Benchmark. The *Multi-Level Academic Survey Test (MAST)* is a timed, norm-referenced, pencil-and-paper test that measures reading comprehension using a time-efficient maze procedure. The test provides norms for students in grade 2 through grade 12, providing a more sensitive measure of growth for older students who perform at the low end of the distribution than a traditional norm-referenced test. The MAST includes a short form, which requires 20 minutes to administer, and a long form, which requires approximately 40 minutes.

Because the measure includes easier items for the lower levels of reading proficiency a reliable distribution can be obtained at these lower levels for older lower-performing students. However, the grade level norms collapse the distribution into single digit percentiles for these lower performing older students. To provide a more sensitive measure of relative distribution, we used the performance of a national sample of grade 9 SWDs to calculate percentile scores for raw scores on the short form of the MAST. We used the short form because we had a more

complete data set for that form of the MAST. This sample included the complete population of grade 9 SWDs (total $n = 134$) at 9 high schools. SWDs are the higher performing students with disabilities who take at least half their coursework in mainstream classes. Although not selected with statistically rigorous consideration of representativeness in mind, the high schools nevertheless included a full range of the types of high schools found in America: 3 urban poverty schools, 3 rural poverty schools, and 3 suburban mixed schools. The schools also sampled a range of geographic areas: 3 were in Kansas, 2 in California, 1 in Oregon, and 3 in Washington state. The percentile scores referenced to the performance of this national sample of 9th grade SWDs provide an indicator of a student's relative rank in this sample.

Reasoning Benchmark. The *New Jersey Test of Reasoning Skills* (Shipman, 1983) is a 50-item pencil-and-paper test of elementary reasoning and inquiry skills. The reliability of the test for students in grades 7 and above is reported in the manual to be .91. The New Jersey Test was originally designed to assess the effectiveness of *Philosophy for Children*, a program for teaching critical thinking. An effort was made in creating the test "to construct a taxonomy in terms of the skills needed to perform the operations in the discipline of logic, both of the formal and informal variety, insofar as these relate to linguistic usage.... The taxonomy appears to be a reasonable representative of the domain, and the items selected for the New Jersey Test appear to be reasonably representative of the taxonomy" (Shipman, 1983, p. 14).

Integrity of Implementation Checklist. The research team designed this observation rubric to evaluate the integrity of the implementation of the considerate instruction on a scale of 0 to 3. All included groups reached at least the minimal criterion of a "2" on this scale. This is the minimal level of implementation fidelity that, based on anecdotal evidence, seems required to achieve significant results.

Experimental Comparisons

Subjects

A sample of 29 SWDs, who were placed for no more than 50% of the day in a special education class, were selected from a high school that served a rural population of high poverty families. To provide a point of reference and a context for evaluating the change in performance of the students with disabilities, we also included 50 at-risk students in the experimental study.

All of the students, both those with disabilities and those considered to be at-risk, had serious reading problems involving an inability to decode fluently or proficiently. Prior to treatment, these students all scored on the MAST reading subscale (short form) at a level lower than the average score achieved by the MAST grade 6 norming sample.

Procedures

Two special education teachers matched the students with disabilities in pairs based on the teachers' perceptions of their performance levels, and their classroom demeanor. After

independently ranking the students, the two teachers met together to resolve discrepancies. A student from each pair was randomly assigned to either the considerate treatment or the control group. The at-risk students were matched in pairs based on their performance on the curriculum-based *Corrective Reading Decoding Program* and assigned to treatment or control group. Five SWDs were taught with the at-risk students. The remaining 24 SWDs were taught in self-contained language arts classes designed specifically for SWDs—one an experimental class the other a control.

The experimental and control general education classes were taught by general education teachers and were designed specifically for students who were at risk of failure to graduate from high school. Both treatments were designed to teach to the standards of the high school exit exam; only the experimental treatment used the considerate instructional design. No more than one student with a disability was present in a general education class.

The experiment continued for one 9-week period on a block schedule. The contrasting experimental-and-control treatment was offered as one-half of a block, so the experimental and control groups received about 40 minutes of the contrasting instruction two and three times a week over a 9 week period. The total amount of instructional time was approximately 15 hours.

The Control Treatment

The control classes received instruction in "The Write Path," which emphasized connections to real life and is part of the AVID program for preparing at-risk students for college. Learning activities involved discussions about real-life problems and extensive opportunities to write about those problems using problem-solving frameworks.

The Experimental Treatment

The experimental group completed ten lessons of the considerate treatment, *Reasoning and Writing Level E* (Engelmann & Grossen, 1999), and the first mastery test. The teacher of the experimental groups was frequently ill and missed about 15 days of school, including days when coaching was offered to ensure that the teacher was delivering the instruction with fidelity. Consequently, the treatment group received approximately 10 hours of instruction from the trained teacher. However, for the ten lessons that were taught, the groups achieved the required mastery levels. (Eighty percent of the group is required to achieve at least 90% correct on the mastery test or the teacher reteaches specific segments of the lessons.) If the additional coaching had been provided as planned, the teacher might have become more efficient, achieving the required mastery levels with less reteaching. Nonetheless, because the group reached the mastery criteria on the first 10-lesson test, we can conclude that the program was implemented with the minimal required level of fidelity for the first 10 lessons.

Baseline Groups

A group of 25 SWDs representing the entire population of SWDs at the rural poverty school were given the MAST in April. Their scores are reported as a baseline comparison for the instructional groups: reading short form mean = 23.4 ($sd = 9.0$); reading long form mean = 34.2 ($sd = 13.6$). An additional group of at-risk students ($n = 37$) at the rural poverty school were also given the MAST. Their scores are used as a baseline for the at-risk students: reading short form mean = 31.8 ($sd = 6.7$), reading long form mean = 47.2 (10.4).

Results

The special education teacher of the self-contained class of SWDs implementing the experimental treatment dropped out of the study before completion. Only the pretest data were collected and those data are included in the data summaries in Table 3 below to see how their initial performance level may have differed from that of students who remained in the study. Their mean scores of the special education groups were significantly lower than those of the at-risk groups on the reading pretest, $t(37) = 2.04, p < .05$. The special education group seemed to lack the pre-skills necessary to learn from the standards-based program, although the SWDs who learned with the at-risk group scored lower initially than the experimental group of SWDs, and made stronger pre-to-post gains than the at-risk students with whom they learned.

Because the pretest scores were equivalent for the experimental and control at-risk general education groups, we used a simple analysis of variance to compare the posttest performance of the groups on reasoning and on reading. Neither difference was significant; however, the differences on the MAST reading comprehension measure (long form) approached significance, $t(1,48) = 1.32$.

Using the average standard deviations of the two treatment groups we calculated effect sizes as a measure of the amount of growth achieved by the SWDs in the respective treatments. The effect sizes on the New Jersey reasoning measure for the SWDs in the considerate treatment were consistently more than double that of the SWDs in the control treatment. On the reading comprehension measure (MAST long form), two of the three SWDs in the control group had negative effect sizes indicating a loss in overall reading competence. In one case the negative effect size was quite large ($-.82$). The one SWD with a positive effect size did not match the positive effect size of the considerate treatment group, nor did his effect size match that of the one SWD who took the reading posttest.

Table 4 displays these effect sizes. Figures 1 and 2 graphically display the slopes of the gains for the groups and for the individual students with disabilities. Visual inspection indicates a consistently steeper slope for the SWDs receiving the considerate instruction.

To evaluate the change in performance of the subjects of the study in terms of relative rank in the national sample of grade 9 SWDs, we converted the MAST short form raw scores to

the percentile scores derived from that national SWD sample. Table 5 displays these scores. Only the scores for the experimental considerate treatment show a gain. The scores for the control group and the SWDs in the control treatment were all negative or zero.

Quasi-Experimental Comparisons

Subjects

Six grade 9 SWDs from the suburban mixed SES public high school participated in the quasi-experimental study. Three SWDs, who were assigned to a special education placement for less than 50% of their instruction, attended the general education language arts classes for at-risk students that were taught using the considerate instructional design. Three attended special education language arts classes designed specifically for special education students.

The at-risk grade 9 students who received considerate instruction and the grade 9 normally achieving general education students were also included in the study. In addition, quasi-experimental comparisons were made with other instructional groups from the other high school settings.

Procedures

At-risk students were placed into one of three types of considerate instruction based on their performance. The treatment groups were as follows:

Considerate Remedial (ConsRem).

Students who received remedial instruction were low readers who began receiving considerate instruction in grade 9 (*Corrective Reading—Decoding*, Engelmann and associates, 1999). These students failed to meet minimum criteria on a decoding test (approximately grade 6 level).

Considerate Standards-based 1 (ConsSB1).

Students who received standards-based instruction were low readers who began receiving considerate instruction in middle school and continued into high school to receive standards-based language arts instruction. The specific instructional program was the *Corrective Reading—Comprehension Level C* program (Engelmann and associates, 1999).

Considerate Standards-based 2 (ConsSB2).

As in the above group, these students began receiving considerate instruction in middle school. However, the specific instructional program these students received was *Reasoning and Writing Level E* (Engelmann & Grossen, 2001). This is the same program used in the considerate treatment in the experimental study. These students had already taken the first standards-based course and, in contrast to the experimental groups, entered the course with a higher pre-skill level of performance.

General Education Regular Curriculum (ControlReg).

The performance of normally achieving students receiving instruction in the regular high school language arts curriculum was compared with the performance of the groups receiving considerate instruction.

Of the three SWDs placed into the considerate treatments, two of them were placed with the at-risk group receiving considerate standards-based 1 instruction. One received instruction with the at-risk group receiving considerate remedial reading instruction. The three SWDs receiving the control treatment received their instruction in special education pullout classes designed specifically for their needs. They were not placed the regular education control group. All of the students in the above groups were from the suburban mixed high school. All groups with pre-and-posttest scores received approximately 25 hours of instruction between those testing occasions

Other Comparison Groups

We also examined the following additional groups of at-risk students receiving considerate instruction in other settings.

1. *Traditional schedule considerate remedial instruction (ConsDailyRem).* A sample of 26 at-risk students from a high poverty urban high school in a Midwestern state, who received the considerate treatment in a traditional schedule of daily instruction for 45 minutes a day, was selected for comparison. All students in grades 9 to 11 who scored on the MAST reading subscale (short form) at a level lower than the average score achieved by the MAST norming sample of grade 6 students were selected for the sample. These 26 students included 17 grade 9 students, 7 grade 10 students, and 2 grade 11 students.

The 26 subjects received considerate instruction in basic reading skills during one 45-minute period per day for 5 months—approximately 50 hours of instruction. The considerate program was *Corrective Reading—Decoding*. Students were grouped into three different levels with students at a similar reading level for instruction. The MAST was administered as a pretest in November and a posttest in April.

2. *Block schedule considerate standards-based instruction (ExpAt-Risk).* The performance of the group receiving considerate instruction in the experimental comparison is included here also for quasi-experimental comparisons. This group received 10 hours of instruction every other day for a 9-week quarter and covered 10 lessons of the standards-based considerate instruction with mastery.

3. *Block schedule control standards-based instruction (ControlAt-Risk).* The performance of the group receiving the control treatment (the AVID program) in the experimental study is also included here for quasi-experimental comparisons. This group received about 15 hours of instruction on a block schedule every other day for 9 weeks.

In addition, the mean scores of the baseline SWD and at-risk groups for the rural poverty school are included in the comparisons.

Results

Group Differences on the SAT-9 and HSEE

To evaluate the effects of the treatments on general language arts outcomes, we collected raw scores on the SAT-9, calculated the means for the various groups, and then converted the mean raw score to a percentile. We used only matched scores so students without a score from the previous year were not included in the analysis. Students who received the considerate remedial instruction were generally new to the district, and did not have SAT-9 scores for the previous year, so the considerate remedial group is not included in this analysis. Similarly, many of the subjects in the control regular education class were also new to the district. Generally, only the students returning to the district who have never been in need of remedial instruction were in the control regular education group.

Table 6 displays the results of these analyses. There was a sharp drop in the reading percentile equivalents from the previous year for all groups. The control regular education group experienced the steepest decline, a loss of 13 percentile points. The considerate instructional groups declined 8 and 10 percentile points.

We checked the consistency of the percentile scores with a larger sample, over 300,000 grade 9 students in California. The percentile for the mean raw score for all grade 9 students tested in reading declined 12 points, while the language score remained the same. This large drop for such a large population raises questions about the representativeness of the norming sample for the SAT-9. With such a large sample the percentile score for the mean raw score should remain at about the 50th percentile from one year to the next, if the norming sample was representative.

The change in language percentile scores varied across the instructional groups of our study. The scores of the considerate standards-based1 group (the initially lower-performing standards-based group) showed an increase of 10 percentile points; the considerate standards-based2 group remained flat; and the scores of the control regular education group dropped 1 percentile point.

To test for differences among the groups in learning gains as measured by the SAT-9 we used an analysis of covariance (ANCOVA) with the mean raw scores from the previous year as the concomitant variable. Table 6 displays the means and standard deviations for matched scores for the groups represented in this analysis. According to the ANCOVA, the groups did not differ significantly in slope or elevation.

In addition, the California High School Exit Exam was administered to all grade 9 students. Table 7 displays the means and standard deviations of the scores for the three groups.

An additional MANCOVA test, using the SAT-9 reading raw score as the concomitant variable, indicated significant differences among groups. A Scheffe' test indicated that the only significant difference was between the considerate standards-based1 treatment and the control regular education treatment on the California High School Exit Exam, favoring the control regular education treatment.

The mean score on the HSEE for all groups was higher than the pass score of 350.

The Performance of Students With Disabilities on the SAT-9 And HSEE

We evaluated the change in performance of the students with disabilities by charting their percentile scores on the SAT-9 reading and language subscales. Table 8 presents these scores. SWDs 1 and 2 were taught in the considerate standards-based 1 treatment, SWD 3 was taught in the considerate remedial treatment, and the control SWDs were taught in the resource room. Figure 3 graphically displays the change in performance of these students as indicated by the percentile scores on the language subscale of the SAT-9. Of the 12 pre-to-post scores for SWDs, only two scores showed a rise from pretest to posttest. Those were for two SWDs in the area of language. Only one SWD received a pass score. That student (SWD 11) was in the control group and also achieved the highest score on the pretest. SWD 7 nearly passed. SWD 7 had been in the considerate treatment for two years prior at the middle school level. By grade 9 he nearly passed the HSEE on the first try.

Performance on the MAST

Percentile equivalents using grade level norms, and grade level equivalents are also reported. Because the pre-and-posttests included a shorter interval than a full year, comparing the pretest to grade 8 norms and the posttest to grade 9 norms, is somewhat unfair. The change in raw scores is displayed to show that some growth occurred. The only group showing a rise in percentile score was the considerate standards-based2 group. We used a MANCOVA to test for differences among the groups, using the pretest as the concomitant variable. The results indicated that the considerate standards-based2 group scored significantly higher on the posttest than the control group ($p < .05$). No other differences were significant.

To determine the relative rank of the various groups and the individual SWDs in the national sample of grade 9 SWDs we used the MAST short form scores. Table 10 displays each SWD and each group's relative rank with the national sample of grade 9 SWDs as a percentile. The group receiving considerate daily remedial instruction (ConsDailyRem) made significantly greater gains than.

Discussion

In this study we faced two major new challenges. We tested whether we could narrow the achievement gap experienced by high school SWDs and at-risk students in the broad domain of language arts. Our measures, therefore, covered a broad domain, rather than a narrow domain as

in previous short duration studies. Second, we tested the most powerful intervention model we know in high-need schools. Consequently, the teachers were not experts in the model working in normally achieving schools, as in our previous studies, but the teachers working and living in that high-need environment every day.

The Rural Poverty Setting

In the rural poverty setting, the experimental comparison of a relatively large number of SWDs was dropped before completion. SWDs were placed in the self-contained class versus the general education class based on their IEP, which could mean that the self-contained students were lower performing. To determine whether the preskills of the self-contained group of SWDs may have been lower, thus making the standards-based intervention too difficult for them, we compared their pretest scores to the rest of the groups. The performance level of these SWDs was lower than that of the other groups. Further research is needed to explain why two SWDs learning in the general education class seemed to do quite well in spite of having less initial reading competence than the group of SWDs who could not manage the material.

The differences in posttest performance of the at-risk groups remaining in the study approached significance on the MAST test of reading comprehension (long form) favoring the considerate treatment. Interestingly, the SWDs in the considerate at-risk classroom seemed to improve at a faster rate than their at-risk colleagues on the New Jersey Test of Reasoning. Unfortunately, only one SWD was present for the MAST reading posttest.

Possibly a longer duration for the study would have produced significant results. We attempted to continue the study into the second semester; however, the circumstances of the high school schedule required extensive changes in each student's program. It was impossible within the current high school framework to extend the intervention.

The Suburban Mixed SES Setting

In contrast to the experimental study in the rural poverty school, where special education students were randomly assigned to the standards-based instruction (*Reasoning and Writing*), students from the suburban setting were placed in instructional groups based on their performance on a placement test. The students placed in the standards-based 2 (ConsSB2) program, *Reasoning and Writing*, were only able to do so because they had taken considerate courses at the middle school level that prepared them for this program. Consequently, the students in the higher group (SB2) were able to achieve scores that were well within passing range of HSEE. Although the mean score of the ConsSB2 group was nonsignificantly lower than the control regular education group, the standard deviation was smaller. The score one standard deviation below the mean for the SB2 group (358) was higher than that of the control group (351). More students in the SB2 group passed the HSEE than in the control group: 52% of the SB1 group passed; 88% passed in the SB2 group; and 81% passed in the control group.

The higher level standards-based program unfortunately contained no SWDs. Two SWDs were included in the SB1 group, and one of those came within 2 points of passing the HSEE. Similarly, one SWD in the control group passed the HSEE by 4 points.

Conclusions

Overall the results of our study indicated that with less than 50 hours of considerate instruction, we can barely begin to narrow the achievement gap when the students begin the solution at grade 9 level. The attempt to drop SWDs into the standards-based considerate instruction, regardless of whether they had the prerequisite skills, failed. Our two groups of at-risk students who began the considerate intervention in middle school and continued into high school, moving into the standards-based considerate programs with the necessary prerequisite skills, seemed closer to attaining our goal.

To begin to close the achievement gap for students who are significantly behind at the grade 9 level will require more time for considerate instruction and / or more intensity in the delivery of the instruction. We noted anecdotally that there was room to increase the intensity of the instruction in most of the classrooms. Many minutes of precious instructional time were wasted, especially in high schools with block schedules.

In addition, high school models that allow more instructional time for accelerated catch up using considerate programs are needed. The rural poverty high school actually has begun a program where students who are significantly behind are pulled out of the regular high school setting and placed in an alternative high school, where a full-day curriculum of considerate instruction is offered until the students reach critical benchmarks. When they reach these benchmarks, they may re-enter the regular high school. Students with disabilities, however, are placed in the regular high school. They, of course, already have a pullout program.

In summarizing what we know about closing the learning gap for students with disabilities, Lyon et al. (2001) suggested that remediation models for older children have been ineffective for two main reasons. "First, the instruction provided through remediation is frequently too little, too general, and too unsystematic. Secondly, even if the instruction were of high quality, it may be too late given that many children are already far behind and less motivated to learn to read" (p. 272). An intensive schedule with considerate instructional programs seems to have the most promise for closing the gap for high school SWDs. We found no shortcuts.

References

- Carnine, D. (March, 1994). The BIG Accommodation program. *Educational Leadership*, 87-88.
- Carnine, D., Caros, J., Crawford, D., Hollenbeck, K., Harniss, M. (1996). Designing effective U.S. history curricula for all students. In J. Brophy (Ed.), *6th Handbook on Research on Teaching*. Advances in Research on Teaching (Vol.6), 207-256. Greenwich, CT: J.A.I. Press.
- Caros, J. (2001). *Emerald Ice 4th-dimension database*. Vancouver, BC: Seira.
- Collins, M., & Carnine, D. (1988). Evaluating the field test revision process by comparing two versions of a reasoning skills CAI program. *Journal of Learning Disabilities*, 21, 375-379.
- Engelmann, S., & Grossen, B. (1999). *Reasoning and writing level E*. Blacklick, OH: Science Research Associates.
- Foorman, B., Francis, D., Beeler, T., Winikates, D., & Fletcher, J. (1997). Early interventions for children with reading problems: Study designs and preliminary findings. *Learning Disabilities*, 8, 63-71.
- Fuches, L. S., Fuches, D., and Kazdan, S. (1999). Effects of peer-assisted learning strategies on high school students with serious reading problems. *Remedial and Special Education*, 20(5), 309-319.
- Grossen, B. (1998). The research base for *Corrective Reading*, SRA. Blacklick, OH: SRA.
- Grossen, B. (2002). The BIG Accommodation Model. *Journal for the Education of Students Placed At Risk*, 7(2), 241-263.
- Grossen, B. (in press). Success of a direction instruction model at a secondary level school with high-risk students. *Reading and Writing Quarterly*.
- Grossen, B., & Carnine, D. (1990). Diagramming a logic strategy: Effects on more difficult problem types and transfer. *Learning Disability Quarterly*, 13, 168-182.
- Grossen, B., Caros, J., Carnine, D., Davis, B., Deshler, D., Schumaker, J., Bulgren, J., Lenz, K., Adams, G., Jantzen, J., & Marquis, J. (2002). Big ideas (plus a little effort) produce big results. *Teaching Exceptional Children*, 34(4), 70-73.
- Grossen, B., Lee, C., & Johnston, D. (1995). A comparison of the effects of direct instruction in reasoning with constructivism on deductive reasoning. In A. Deffenbaugh, G. Sugai, & G. Tindal (Eds.), *The Oregon Conference 1995 Monograph* (pp. 253-273). Eugene, OR: College of Education, University of Oregon.
- Hanushek, E., Kain, J., & Riukin, S. (1998). *Does special education raise academic achievement for students with disabilities?* National Bureau of Economic Research working paper no. 6469.

- Kameenui, E., & Carnine, D. (2001). *Effective teaching strategies that accommodate diverse learners*. Columbus, OH: Merrill.
- Lyon, R., Fletcher, J., Shaywitz, S., Shaywitz, B., Torgesen, J., Wood, F., Schulte, A., & Olson, R. (2001). *Rethinking learning disabilities*. Progressive Policy Institute, Fordham Foundation.
- Moody, s., Vaughn, S., Hughes, M. & Fischer, M. (2000). Reading instruction in the resource room: Set up for failure. *Exceptional Children*, 16, 305-316.
- Schumm, J., Moody, S., & Vaughn, S. (2000). Grouping for reading instruction: Does one size fit all? *Journal of Learning Disabilities*, 5_477-488.
- Torgesen, J., Alexander, A., Wagner, R., Rashotte, C., Voeller, K., Conway, T., & Rose, E. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities*, 34, 33-58.
- Vaughn, S., Moody, S., & Schumm, J. (1998). Broken promises: Reading instruction in the resource room. *Exceptional Children*, 64, 211-225.

Table 1.

The Contrast Between Instruction with Accommodations for Diverse Learners and Traditional Instruction

Six Principles of Accommodation for Diverse Learners	Traditional Instruction
<u>Big Ideas</u> , concepts and principles that facilitate the most efficient and broad acquisition of knowledge across a range of examples, are presented. Big ideas make it possible for students to learn the most and learn it as efficiently as possible, because "small" ideas can often be best understood in relationship to larger, "umbrella concepts."	A barrage of unrelated facts and details are presented. The links between concepts are obscured.
<u>Conspicuous Strategies</u> made up of specific steps that lead to solving complex problems are taught.	Strategies are seldom taught.
<u>Background Knowledge</u> is pre-taught.	Important prerequisite learning is often not evaluated nor taught.
<u>Mediated Scaffolding</u> provides personal guidance, assistance, and support that gradually fades as students become more proficient and independent.	Little direction or provision for scaffolding the progression of learning toward greater independence is provided.
<u>Judicious Review</u> requires students to draw upon and apply previously learned knowledge over time.	Review is often minimal.
<u>Strategic Integration</u> blends new knowledge with old knowledge to build bigger big ideas.	Spiraling of topics does not carefully integrate units.

Table 2
Demographic Characteristics for Three High School Settings

Characteristic	Rural Poverty	Suburban Mixed	Urban Poverty
Participants eligible for free or reduced lunch	57%	24%	44%
African-Americans	5%	28%	17%
Latinos	53%	32%	12%
Asian	1%	14%	14%
Total school enrollment	498	635	2181
Location	California	California	Midwest
Scheduling format	Block (groups meet every other day)	Block (groups meet every other day)	Traditional (groups meet every day)

Table 3.
Descriptive Statistics of Performance of Students with Disabilities (SWD) and the Comparison Groups on the New Jersey Test of Reasoning and on the Multi-Level Academic Survey Test

Group	n	Pre-Test		Post-Test		Rank in Group
		Reasoning ¹	Reading ²	Reasoning	Reading	
Considerate Language Arts Instruction						
SWD 1		12	32	20	41	19->8
SWD 2		14	--	23	--	13->4
Exp Sped	11	13.4 (3.8)	36.5 (10.0)			
Exp At-Risk	28	13.9 (4.7)	46.3 (10.7)	16.0(4.7)	49 (8.3)	
Control Language Arts Instruction						
SWD 3		8	32	10	29	21->19
SWD 4		11	35	14	27	13->14
SWD 5		14	48	17	50	9->4
Ctrl Sped	13	13.6 (3.7)	36.0 (11.1)	--	--	
Ctrl At-Risk	22	13.7 (4.7)	46.6 (8.9)	15 (4.9)	45.4 (11.0)	
Baseline SWD	25				34.2 (12.6)	
Baseline At-Risk	37				47.2 (10.4)	

¹Reasoning= New Jersey Test of Reasoning (Shipman, 1983)

²Reading= MAST

Table 4

Effect Sizes for Pre- to Post-test Achievement Gain for Students with Disabilities and the Comparison Groups on the New Jersey Test of Reasoning and on the MAST Reading Test (long form)

Students	Effect Size	
	Reasoning (sd = 4.7)	Reading (sd = 9.8)
Experimental at-risk group	.45	.28
Exp SWD 1	1.7	.92
Exp SWD 2	1.9	--
Control at-risk group	.28	-.12
Control SWD 3	.43	-.31
Control SWD 4	.64	-.82
Control SWD 5	.64	.20

Note: Standard deviations for the denominator were calculated by averaging the standard deviations for the two groups on the pretest.

Table 5.

Performance on the MAST Short Form Reported in Mean Raw Scores with Standard Deviations, and Percentile Scores Derived from a National Sample of SWDs.

	Pretest		Posttest		Change in %tile Rank
	Mean (sd)	%tile	Mean(sd)	%tile	
Exp SWD 1	22	26	29	41	15
Exp SWD Group (n = 11)	26.8 (8.2)	36	--	--	--
Exp At-Risk Group (n=28)	31.5 (5.7)	49	33.6 (3.8)	54	5
Control					
Control SWD 3	28	38	23	28	- 10
Control SWD 4	27	36	21	25	- 2
Control SWD 5	36	65	36	65	0
Control SWD Group (n = 13)	25.3 (6.0)	33	--	--	--
Control At-Risk (n=22)	32.8 (4.2)	52	31.4 (6.4)	49	- 3
Baseline SWD (n =25)	--	--	23.4 (9.0)	29	--
Baseline At-Risk (n=37)	--	--	31.8 (6.7)	49	--

Table 6

Percentiles for the Mean Raw Matched Scores on the Reading and Language Subscales of the SAT-9 for Instructional Groups in the Suburban Mixed SES High School

Treatment	N	Percentile of the SAT-9 Mean Raw Matched Score			
		Reading		Language	
		Pre	Post	Pre	Post
ConsSB1	59	35	25	31	41
ConsSB2	23	51	43	60	60
Control Reg	103	54	41	58	57

Table 7

Mean Raw Scores and Standard Deviations for the Performance of the Treatment Groups on the SAT-9

Group	N	Mean Raw Scores (SD)				
		Reading		Language		HSEE
		Pre	Post	Pre	Post	Post
ConsSB1	59	51 (11)	45 (14)	27 (7)	25 (9)	355 (29)
ConsSB2	23	58 (8)	56 (9)	34 (4)	31 (5)	376 (18)
Control Reg	103	60 (12)	55 (14)	33 (8)	30 (10)	383 (32)

Table 8
Percentile Scores of Students with Disabilities on the SAT-9 Reading and Language Arts

Group	SAT-9 Reading		SAT-9 Language		HSEE
	Pre	Post	Pre	Post	Post
Experimental					
Cons SWD 6	13	6	20	38	305
Cons SWD 7	14	14	41	28	348
Cons SWD 8	19	4	11	14	317
Control					
Ctrl SWD 9	9	3	14	11	307
Ctrl SWD 10	10	4	5	1	300
Ctrl SWD 11	26	14	29	23	354

Table 9

Performance of the Treatment Groups on the Multi-Level Academic Survey Test (Long Form) in Mean Raw Score, Standard Deviation, Percentile Equivalent, and Grade Equivalent.

Group	N	Pretest		Posttest		Raw Score
		Mean (SD)	PR ¹ / GE	Mean (SD)	PR ² / GE	Change
Cons SWD 6		35	5 / 3.1	30	2 / 2.7	- 5
Cons SWD 7		42	6 / 4.7	43	7 / 4.9	+ 1
Cons SWD 8		52	28 / 6.7	53	24 / 7.1	+ 1
Cons SB1	28	54.3 (6.3)	34 / 7.3	55.1 (10.4)	32 / 7.4	+ 0.8
Cons SB2*	19	56.7 (7.6)	44 / 7.8	59.1 (6.6)	50 / 9.0	+ 2.4
Control At-Risk*	22	46.6 (8.9)	13 / 5.5	45.4 (11.0)	9 / 5.1	-1.2
Control Rural SWD	25	--	--	34.2 (12.6)	3 / 2.9	--

¹ Percentile value using end of grade 8 norms

² Percentile value using end of grade 9 norms

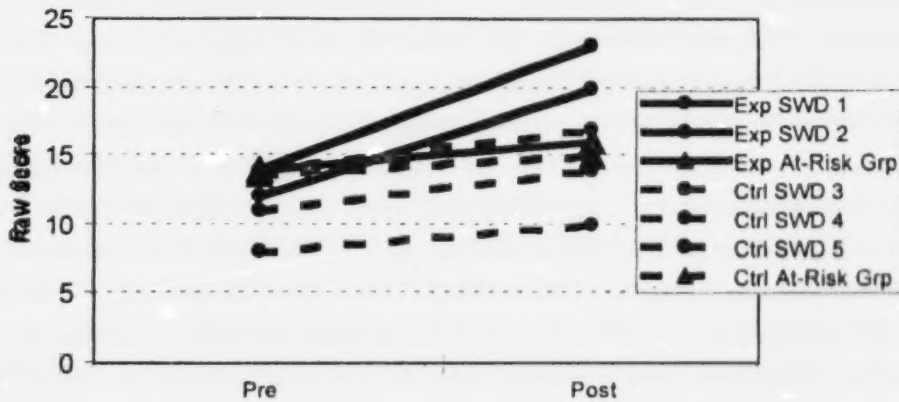
*Difference between these two groups significant at a $p < .05$ level.

Note: Exp SWD 6 received the ConsRem treatment. Exp SWDs 7 and 8 received the ConsSB1 treatment.

Table 10.

Descriptive Statistics of the Performance of Quasi-Experimental Comparison Groups on the MAST (Short Form) and of Individual SWDs with Percentile Scores Derived from a National Sample.

		Pretest	SWD	Posttest	SWD	%ile
	N	Mean (SD)	%ile	Mean (SD)	%ile	Change
<u>Experimental on Daily Schedule</u>						
ConsDailyRem	26	24.4 (8.9)	31	29.3 (9.5)	42	+ 11
<u>Considerate on Block Schedule</u>						
Cons SWD 6	1	21	25	24	30	+ 5
Cons SWD 7	1	34	55	33	53	- 2
Cons SWD 8	1	32	51	37	69	+ 18
ConsBlockSB1	23	33.6 (6.1)	53	36.0 (4.5)	63	+ 10
ConsBlockSB2	19	36.8 (3.5)	68	38.2 (1.9)	75	+ 7
ExpAtRisk	28	31.5 (5.7)	49	33.6 (3.8)	54	+ 5
<u>Control Groups on Block Schedule</u>						
ControlAtRisk	22	32.8 (4.2)	52	31.4 (6.4)	49	- 3
BaselineRuralSWDs	25	--	--	23.4 (9.0)	30	--
BaselineRuralAtRisk	37	--	--	31.8 (6.7)	49	--



Note: Groups are indicated with a triangular-shaped point. Individual students with disabilities are indicated with a circular point.

Figure 1. Graphic display of the pre-to-post gain for comparison groups and for students with disabilities on the New Jersey Test of Reasoning.

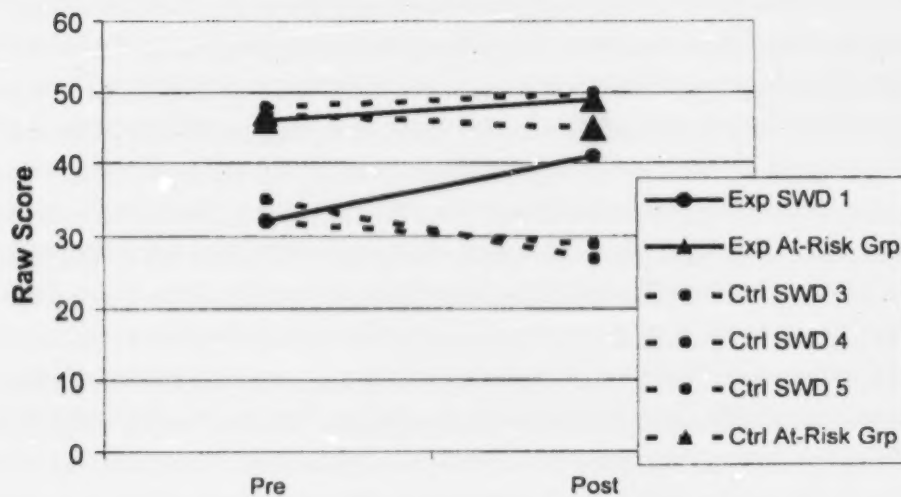


Figure 2. Graphic display of pre-to-post gain for comparison groups and for students with disabilities on the Multi-Level Academic Survey Test.

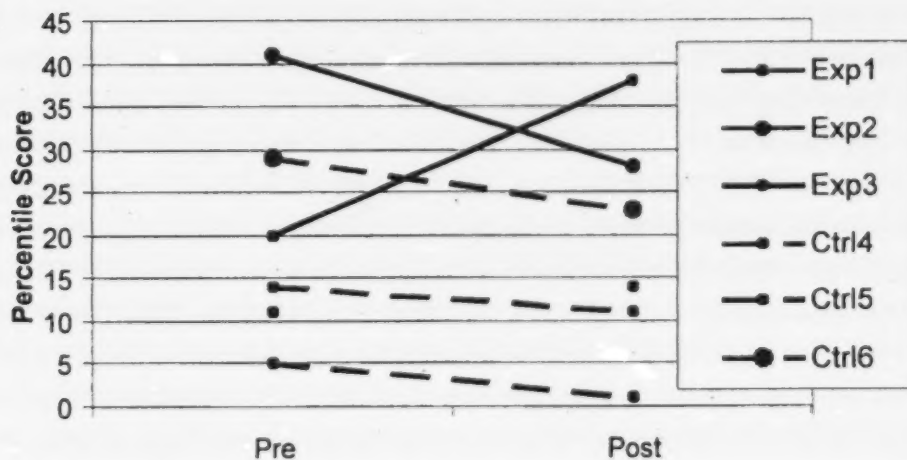


Figure 3. Graphic display of change in performance of students with disabilities on standards-based measures (SAT-9).



U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)



NOTICE

Reproduction Basis

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☒ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").